

Rational decomposition of dense hypergraphs and some related eigenvalue estimates

Peter Dukes

Department of Mathematics and Statistics
University of Victoria, Victoria, BC V8W3R4, Canada
 dukes@uvic.ca

Abstract

We consider the problem of decomposing some family of t -subsets, or t -uniform hypergraph G , into copies of another, say H , with nonnegative rational weights. For fixed H on k vertices, we show that this is always possible for all G having sufficiently many vertices and density at least $1 - C(t)k^{-2t}$. In particular, for the case $t = 2$, all large graphs with density at least $1 - 2k^{-4}$ admit a rational decomposition into cliques K_k . The proof relies on estimates of certain eigenvalues in the Johnson scheme. The concluding section discusses some applications to design theory and statistics, as well as some relevant open problems.

Keywords: graph decomposition, fractional decomposition, t -graph, t -design, Johnson scheme, eigenvalue perturbation

2000 MSC: 05B05, 05C70

1. Preliminaries

Let t be a positive integer. The set of all t -element subsets of a set X is written $\binom{X}{t}$. By a (rational) t -vector on X , we mean a function $f \in \mathbb{Q}^{\binom{X}{t}}$.

A t -uniform hypergraph, or simply t -graph is a triple $H = (X, E, \iota)$, where

- X is a set of *points* or *vertices*,
- E is a set of *edges*, and
- $\iota \subset X \times E$ is an *incidence* relation such that every edge is incident with precisely t different vertices.

Edges are usually identified with the set of incident vertices, dispensing with ι . However, the definition above permits ‘multiple edges’. If there are no multiple edges, then H is said to be *simple*. Unless otherwise specified, all t -graphs will be assumed simple, and $E \subseteq \binom{X}{t}$. With this understanding, we may conveniently identify t -graphs with $(0, 1)$ t -vectors.

A t -graph H' with vertex set X' and edge set E' is a *subgraph* of H if $X' \subseteq X$ and $E' \subseteq E$. The corresponding t -vectors satisfy $f' \leq f|_{\binom{X'}{t}}$.

Ordinary graphs are 2-graphs; note however that the definition does not allow ‘loops’.

Consider a large t -graph G on vertex set V , $|V| = v$. For $0 \leq s \leq t$, the *degree* in G of an s -subset S of vertices is the number of edges of G which contain S . The minimum degree over all s -subsets is denoted $\delta_s(H)$ and we say that G is $(1 - \epsilon)$ -dense if $\delta_{t-1}(G) \geq (1 - \epsilon)(v - t + 1)$. In other words, a t -graph is $(1 - \epsilon)$ -dense if, given any $t - 1$ points, the probability that another point fails to induce an edge is at most ϵ .

The *complete* t -graph or *clique* on V corresponds to the set system $\binom{V}{t}$ and, equivalently, the constant t -vector with every coordinate equal to 1. The standard graph-theoretic notation is K_v^t , where the superscript is normally omitted if $t = 2$, or if it is otherwise understood. Of course, complete t -graphs are 1-dense.

Suppose G and H are t -graphs, as above, with respective vertex sets V and X . A *fractional* or *rational decomposition* of G into copies of H is a set of pairs (H_i, w_i) , where

- each H_i is a subgraph of G isomorphic to H ;
- w_i are positive weights such that, for every edge T of G ,

$$\sum_{i: T \in H_i} w_i = 1.$$

To be clear, $T \in H_i$ means that T is an edge of H_i .

There is no loss in generality in assuming the w_i are rational numbers. Note that if the w_i are integers, the result is an ordinary edge-decomposition. Although we do not need the notation very frequently, a reasonable abbreviation is $H \trianglelefteq_{\mathbb{Q}} G$ for rational decomposition and $H \trianglelefteq G$ for ordinary decomposition.

Alternative descriptions are possible. For instance, if H has vertex set X , a rational decomposition can be viewed as a nonnegative formal linear combination of injections X into V , say $\sigma \in \mathbb{Q}_{\geq 0}[X \hookrightarrow V]$, so that $\sigma H = G$.

A (signed) linear combination of injections $\sigma \in \mathbb{Q}[X \hookrightarrow V]$ is not enough, as the following example shows.

Example 1.1. Here $t = 2$. Let $G = C_5$ be the 5-cycle 12345 on $V = \{1, 2, 3, 4, 5\}$, and let $H = K_3$ on a three element set X . Then combining ‘positive’ copies of H on 123, 145 plus a ‘negative’ copy of H on 134 yields a 2-vector G' with pairs $\{1, 2\}$, $\{2, 3\}$, $\{4, 5\}$, $\{1, 5\}$ having weight 1, pair $\{3, 4\}$ having weight -1 , and all other pairs having weight 0. So the five cyclic shifts of G' combine to yield $3G$ (the 5-cycle with every edge tripled). Therefore, there exists $\sigma \in \mathbb{Q}[X \hookrightarrow V]$ with $\sigma H = G$. However, since H is not a subgraph of G , it is clear that there is no such $\sigma \in \mathbb{Q}_{\geq 0}[X \hookrightarrow V]$.

For ordinary graphs G and H , another equivalent formulation arises from the adjacency matrices A_G and A_H . It is easy to see that $H \trianglelefteq_{\mathbb{Q}} G$ (respectively $H \trianglelefteq G$) is equivalent to a decomposition

$$A_G = \sum w_i Q_i^{\top} A_H Q_i,$$

where Q_i are $|X| \times |V|$ $(0, 1)$ ‘injection’ matrices having row sum 1, and w_i are positive rationals (integers).

The following facts are evident from the definitions.

Lemma 1.2. (a) *Both $\trianglelefteq_{\mathbb{Q}}$ and \trianglelefteq are transitive on t -graphs.*

(b) *If H is a t -graph with $p \leq v$ vertices and $q > 0$ edges, then $H \trianglelefteq_{\mathbb{Q}} K_v^t$.*

Remark. Part (a) is quite clear. For (b), it is enough to take each labeled subgraph of H in the complete graph with weight $\binom{v}{t}/qp!\binom{v}{p}$.

Obviously, for $H \trianglelefteq_{\mathbb{Q}} G$, it is necessary that H be a subgraph of G . In fact, every $t-1$ elements of G must belong to enough copies of H to exhaust the degree at that vertex. For instance, large balanced complete bipartite 2-graphs G are nearly $\frac{1}{2}$ -dense but triangle-free. Edges can be thrown in until G becomes nearly $\frac{3}{4}$ -dense and still admit no decomposition. Actually, not much more is known about the density of G failing to admit $H \trianglelefteq_{\mathbb{Q}} G$ apart from this kind of counting analysis.

In this paper, we prove the following existence result on rational decompositions of dense hypergraphs.

Theorem 1.3. *For integers $k \geq t \geq 2$, there exists $v_0(t, k)$ and $C = C(t)$ such that, for $v > v_0$ and $\epsilon < Ck^{-2t}$, any $(1 - \epsilon)$ -dense t -graph G on v vertices admits a rational decomposition into copies of K_k .*

By Lemma 1.2, the same result holds for any t -graph H on k vertices replacing K_k .

In [12], Yuster proved the same result for $\epsilon \approx 6^{-kt}$, although it was admitted that small improvements may be possible. Probabilistic and combinatorial arguments were central. A better result was obtained for ordinary graphs, proved in [11] for $\epsilon = 1/9k^{10}$.

Here, the improvement from Theorem 1.3 is substantial, with a qualitative weakening on the density requirement for general t , and a bound much closer to the density condition for ordinary graphs. Our new upper bound on ϵ is actually $\frac{1}{2} \binom{k}{t}^{-2}$, and (again) insignificant improvements may be possible from the present proof technique.

Our proof of Theorem 1.3 is, at least in principle, constructive. For each edge in G , consider the family of all k -subsets which cover it and induce a clique K_k^t in G . We actually prove the existence of a nonnegative rational combination of these families which gives G . This is done by finding non-negative solutions to a certain linear system, and the problem then reduces to some elementary linear algebra. The outline of the argument is presented in more detail in Sections 2 and 3. The technicalities amount to estimating certain eigenvalues using the theory of association schemes. These details are covered in Sections 4 and 5. To conclude, some applications and further directions are considered in Section 6.

2. Coverage and linear systems

Let V be a v -set, and suppose that $k \geq 2t$. A set system $\mathcal{F} \subseteq \binom{V}{k}$ is said to *cover* $T \in \binom{V}{t}$ exactly λ times if $T \subset K$ for exactly λ elements $K \in \mathcal{F}$. Alternatively, \mathcal{F} is a k -vector and its *coverage* is a t -vector \mathcal{F}^t with

$$\mathcal{F}^t(T) = \sum_{K \supset T} \mathcal{F}(K).$$

In context, we may suppress the superscript t , and instead write $\mathcal{F}(T)$ for the coverage of T by \mathcal{F} .

Now, let $\mathfrak{X} = \binom{V}{k}$ and consider the family $\mathfrak{X}[U]$ of all $\binom{v-t}{k-t}$ k -subsets of V which contain a given $U \in \binom{V}{t}$. Then

$$\mathfrak{X}[U](T) = \binom{v - |T \cup U|}{k - |T \cup U|},$$

since this counts the number of k -subsets containing both T and U . Therefore, we may write

$$\mathfrak{X}[U](T) = \xi_{|T \setminus U|},$$

where

$$\xi_i = \binom{v-t-i}{k-t-i} = \frac{v^{k-t-i}}{(k-t-i)!} + o(v^{k-t-i}).$$

for $i = 0, 1, \dots, t$. This kind of estimation on the orders of binomial coefficients occurs frequently in what follows.

Let $n = \binom{v}{t}$ and identify \mathbb{Q}^n with $\mathbb{Q}^{\binom{V}{t}}$. Define the $n \times n$ matrix M by

$$M(T, U) = \xi_{|T \setminus U|} = \mathfrak{X}[U](T),$$

for $T, U \in \binom{V}{t}$. Clearly, $M^\top = M$. The constant column (row) sum of M is

$$\begin{aligned} \sum_T \xi_{|T \setminus U|} &= \sum_{i=0}^t \xi_i \binom{v-t}{i} \binom{t}{i} = \binom{v-t}{k-t} \binom{k}{t} \\ &= \frac{v^{k-t}}{(k-t)!} \binom{k}{t} + o(v^{k-t}). \end{aligned} \tag{2.1}$$

Observe that (2.1) simply counts the number of k -subsets intersecting a given k -subset in exactly t points.

Although we do not make explicit use of the abundant additional symmetry in M , it is worth noting that the symmetric group \mathcal{S}_V induces an action on \mathfrak{X} which stabilizes M .

At this point, we note that a nonnegative solution \mathbf{x} to $M\mathbf{x} = \mathbf{1}$ induces a rational decomposition $K_k^t \leq_{\mathbb{Q}} K_v^t$. Simply take each $\mathfrak{X}[U]$ with weight $\mathbf{x}(U)$, and the total coverage is

$$\sum_U \mathbf{x}(U) M(T, U) = (M\mathbf{x})(T) = 1$$

on each t -set T . Indeed, $\mathbf{1}$ is an eigenvector of M , and so the unique such \mathbf{x} simply has the reciprocal of (2.1) in each coordinate.

Decomposing a non-complete t -graph G is not so easy. We must restrict our attention to k -subsets that cover only those edges present in G .

To this end, define $\mathfrak{X}|_G$ as the family of all k -subsets which induce a clique in G . In other words, $K \in \mathfrak{X}|_G$ if and only if

- $K \subseteq V$ with $|K| = k$, and
- $T \subset K$ with $|T| = t$ implies T is an edge of G .

Note that $\mathfrak{X}|_G$ is nonempty when G is sufficiently dense.

Now consider $\mathfrak{X}|_G[U]$, the family of all k -subsets on V which contain U and also induce a clique in G . Define the $|G| \times |G|$ matrix \widehat{M} , with rows and columns indexed by edges of G , by

$$\widehat{M}(T, U) = \mathfrak{X}|_G[U](T).$$

Again, \widehat{M} is symmetric, since its (T, U) -entry just counts the number of k -subsets containing T, U , and no non-edges of G . And, most importantly, a nonnegative solution \mathbf{x} to

$$\widehat{M}\mathbf{x} = \mathbf{1}, \tag{2.2}$$

if it exists, yields a rational decomposition of $K_k^t \trianglelefteq_{\mathbb{Q}} G$. Just as in the easy case of complete t -graphs above, each $\mathfrak{X}|_G[U]$ is taken with multiplicity $\mathbf{x}(U)$ to obtain coverage 1 on edges T of G . By construction, the coverage is also zero on non-edges of G .

The basic theme of this article may be summarized as follows: for dense G , our matrix \widehat{M} is a small perturbation of the principal submatrix $M|_G$ of M , restricted to edges of G . This perturbation will be estimated carefully in the next section; however, the relevant lemma in terms of coverages is given here.

Lemma 2.1. *Suppose G is a $(1 - \epsilon)$ -dense simple t -graph.*

- (a) *Given an edge T and i with $0 \leq i \leq t$, there are at least*

$$\binom{t}{i} \binom{v}{i} \left[1 - \binom{t+i}{i} \epsilon + o(1) \right]$$

edges U such that $|T \setminus U| = i$ and $T \cup U$ induces a clique in G .

(b) If T and U are edges of G with $|T \setminus U| = i$ and such that $T \cup U$ induces a clique in G , then there are at least

$$\binom{v-t-i}{k-t-i} \left[1 - \left(\binom{k}{t} - \binom{t+i}{i} \right) \epsilon + o(1) \right]$$

k -subsets containing $T \cup U$ and inducing a clique in G .

Proof. Let J be a set of $j \geq t$ points which induce a clique K_j^t in G . The number of ways to choose a point x in $V \setminus J$ so that $J \cup \{x\}$ also induces a clique is at least $v - j - \binom{j}{t-1}z$, where z is an upper bound on the number of non-edges incident with each $(t-1)$ -subset. With $z = \epsilon(v-t+1)$, and applying induction, the number of ways to extend T to a clique induced by $T \cup U$, of size $t+i$, is at least

$$\frac{1}{i!} \prod_{t \leq j < t+i} \left[v \left(1 - \binom{j}{t-1} \epsilon \right) - O(1) \right].$$

Note the $O(1)$ term depends on t and ϵ but not on v . We now expand the dominant term of the product and invoke the inequality

$$\prod_j (1 - a_j) \geq 1 - \sum_j a_j.$$

Using an identity on the resulting sum of binomial coefficients $\binom{j}{t-1}$, one has the number of such extensions at least

$$\frac{v^i}{i!} \left[1 - \binom{t+i}{i} \epsilon \right] + o(v^i).$$

Finally, in choosing an edge U (not merely an extension of T), we are free to pick any $t-i$ points in T . This proves (a).

Similarly, the number of ways to extend a clique on $T \cup U$ to a clique on k points is at least

$$\frac{1}{(k-t-i)!} \prod_{t+i \leq j < k} \left[v \left(1 - \binom{j}{t-1} \epsilon \right) - O(1) \right],$$

or, after expansion and identities,

$$\frac{v^{k-t-i}}{(k-t-i)!} \left[1 - \left(\binom{k}{t} - \binom{t+i}{i} \right) \epsilon \right] + o(v^{k-t-i}).$$

This proves (b). □

Remarks: Lemma 2.1(a) essentially asserts that ‘most’ entries of \widehat{M} are nonzero, while part (b) asserts that those nonzero entries are close to those of M .

3. Proof of the main theorem

Our proof relies on a couple of easy facts from linear algebra. Recall that the matrix norm $\|\cdot\|_\infty$ is induced from the same (max) norm on vectors. We have $\|A\|_\infty$ equal to the maximum absolute row sum of A . We note below that small perturbations in this norm (actually, in any induced norm) do not destroy positive definiteness.

Lemma 3.1. *Suppose A and ΔA are Hermitian matrices such that every eigenvalue of A is greater than $\|\Delta A\|_\infty$. Then $A + \Delta A$ is positive definite.*

Proof. This follows easily since the spectral radius (i.e. maximum eigenvalue) of ΔA satisfies

$$\rho(\Delta A) \leq \|\Delta A\|_\infty.$$

□

We will momentarily invoke this fact with $A = M|_G$ and $\Delta A = \Delta M := \widehat{M} - M|_G$.

First though, recall Cramer’s rule from college linear algebra. For non-singular A , the system $A\mathbf{x} = \mathbf{b}$ has a solution given by

$$x_i = \frac{\det(A_i)}{\det(A)},$$

where A_i denotes the matrix A with its i th column substituted for \mathbf{b} .

Taken together, we conclude that the system (2.2) has a positive solution \mathbf{x} provided the least eigenvalues of both M and M_1 exceed $\|\Delta M\|_\infty$. Note that we may restrict attention to a single M_1 due to invariance of M under the action of \mathcal{S}_V .

A careful calculation of the eigenvalues of M and M_1 is left for the next 2 sections; however, we summarize the important results here.

Theorem 3.2. *Asymptotically in v , the least eigenvalue of M is*

$$\theta_t = \binom{v-t}{k-t} + o(v^{k-t}),$$

and the least eigenvalue of M_1 is at least $\frac{1}{2}\theta_t$.

Of course, the same lower bounds on eigenvalues remain true for the principal submatrices restricted to rows and columns of M indexed by edges of G .

Now, it remains to estimate the maximum absolute row sum of ΔM .

Proposition 3.3. *Let G be a $(1 - \epsilon)$ -dense simple t -graph, and define ΔM as above. For small ϵ , and asymptotically in v ,*

$$\|\Delta M\|_\infty < \binom{v-t}{k-t} \binom{k}{t}^2 \epsilon + o(v^{k-t}). \quad (3.1)$$

Proof. Let $a(i)$ and $b(i)$ denote the expressions given in the statement of Lemma 2.1, parts (a) and (b), respectively. In row T and columns U with $|T \setminus U| = i$, there are at least $a(i)$ entries where \widehat{M} is nonzero due to $T \cup U$ inducing a clique. That is, there are at most $\binom{v-t}{i} \binom{t}{i} - a(i)$ such entries which vanish in \widehat{M} .

When $T \cup U$ does induce a clique, we have $\widehat{M}(T, U) \geq b(i)$ and $M(T, U) = \xi_i$. That is, ΔM is at most of order $\xi_i - b(i)$ in these entries.

Taken together,

$$\begin{aligned} \|\Delta M\|_\infty &< \sum_{i=0}^t \left[\binom{v-t}{i} \binom{t}{i} - a(i) \right] \xi_i + \binom{v-t}{i} \binom{t}{i} (\xi_i - b(i)) \\ &= \epsilon \sum_{i=0}^t \binom{v-t}{i} \binom{t}{i} \binom{k}{t} \xi_i + o(v^{k-t}). \end{aligned}$$

After invoking (2.1), we obtain the desired bound (3.1). \square

By Lemma 3.1, Theorem 3.2 and Proposition 3.3, the vector $\widehat{M}^{-1} \mathbf{1}$ is (asymptotically in v) entrywise positive for

$$\epsilon < \frac{1}{2} \binom{k}{t}^{-2}.$$

Therefore, we have an induced rational decomposition of G into copies of K_k^t .

We should note that there may be a hope of positive solutions to (2.2) for some (possibly all) graphs G even if this worst-case bound for ϵ were exceeded.

It is worth comparing this method with a similar approach used in [3]. That article considers only $t = 2$ (ordinary graphs) in the special case when G , the graph being decomposed, is circulant. Rather than considering eigenvalues of M and M_1 , a useful lemma on positive determinants was cited from [1]. It would be interesting to see the same method applied for higher t .

Also, it is probably possible to avoid using Cramer's rule and instead analyze the conditioning number $\kappa(M)$. However, this is not likely to yield any substantially better bounds on ϵ .

It now remains to prove Theorem 3.2, and this is the subject of the next two sections.

4. The Johnson scheme and eigenvalue estimates for M

For our purposes, a k -class association scheme on a set \mathfrak{X} consists of $k+1$ nonempty symmetric binary relations R_0, \dots, R_k which partition $\mathfrak{X} \times \mathfrak{X}$, such that

- R_0 is the identity relation, and
- for any $x, y \in \mathfrak{X}$ with $(x, y) \in R_h$, the number of $z \in \mathfrak{X}$ such that $(x, z) \in R_i$ and $(z, y) \in R_j$ is the *structure constant* p_{ij}^h depending only on h, i, j .

Let $|\mathfrak{X}| = n$. For $i = 0, \dots, k$, define the $n \times n$ *adjacency matrix* A_i , indexed by entries of \mathfrak{X} , to have (x, y) -entry equal to 1 if $(x, y) \in R_i$, and 0 otherwise. It is said that x and y are i th *associates* when $(x, y) \in R_i$.

By definition of the structure constants, $A_i A_j = \sum_h p_{ij}^h A_h$. In this way, the adjacency matrices span not only a subspace of the $n \times n$ matrices, but a matrix algebra called the *Bose-Mesner algebra*.

Interestingly, the adjacency matrices are orthogonal idempotents with respect to entrywise multiplication, and

$$A_0 + \dots + A_k = J,$$

the all ones matrix. From spectral theory, the Bose-Mesner algebra also has a basis of orthogonal idempotents E_0, \dots, E_k with respect to ordinary matrix multiplication, and such that

$$E_0 + \dots + E_k = I.$$

A convention is adopted so that $E_0 = \frac{1}{n}J$, which must be one of these idempotents.

For more on the theory of association schemes, the reader is directed to Chapter 30 of [7] for a nice introduction or to Chris Godsil's notes [4] for a very comprehensive reference.

The *Johnson scheme* $J(t, v)$ has as elements $\binom{V}{t}$, where $S, T \in \binom{V}{t}$ are declared to be i th associates if and only if $|S \cap T| = t - i$.

The adjacency matrices and (a certain ordering of) the orthogonal idempotents are related via

$$A_i = \sum_{j=0}^t P_{ij} E_j, \quad (4.1)$$

where $P = [P_{ij}]$ is the *first eigenmatrix*. For $J(t, v)$, its entries are given by

$$P_{ij} = \sum_{s=0}^i (-1)^{i-s} \binom{t-s}{i-s} \binom{t-j}{s} \binom{v-t+s-j}{s}. \quad (4.2)$$

The expression (4.2) is a polynomial of degree $2i$ in j . It is a relative of the family of *Hahn polynomials*. From (4.1), we have

$$M = \sum_{i=0}^t \xi_i A_i = \sum_{j=0}^t \theta_j E_j,$$

where

$$\theta_j = \sum_{i=0}^t \xi_i P_{ij}. \quad (4.3)$$

Since the E_j are orthogonal idempotents, it follows that the eigenvalues of M are θ_j , having multiplicity

$$m_j = \text{rank}(E_j) = \binom{v}{j} - \binom{v}{j-1}.$$

Of course, columns of the E_j are eigenvectors for θ_j .

An easy calculation with convolution identities gives the closed form

$$\theta_0 = \sum_{i=0}^t \xi_i \binom{t}{i} \binom{v-t}{i} = \binom{v-t}{k-t} \binom{k}{t}.$$

This is simply the row sum of M , or (2.1). The other eigenvalues are more complicated, but for our purposes an estimation will suffice.

Proposition 4.1. *The eigenvalues of M are θ_j , each of multiplicity $m_j = \binom{v}{j} - \binom{v}{j-1}$. For sufficiently large v , the θ_j are distinct reals of order v^{k-t} .*

Proof. Computing directly from (4.2) and (4.3),

$$\begin{aligned}\theta_j &= \sum_{i=0}^t \xi_i P_{ij} \\ &= \sum_{i=0}^t \binom{v-t-i}{k-t-i} \sum_{s=0}^i (-1)^{i-s} \binom{t-s}{i-s} \binom{t-j}{s} \binom{v-t+s-j}{s}.\end{aligned}$$

Now separating the $s = i$ term,

$$\begin{aligned}\theta_j &= \sum_{i=0}^{t-j} \binom{v-t-i}{k-t-i} \binom{v-t+i-j}{i} \binom{t-j}{i} + o(v^{k-t}) \\ &= \frac{1}{(k-t)!} \left[\sum_{i=0}^{t-j} \binom{k-t}{i} \binom{t-j}{i} \right] v^{k-t} + o(v^{k-t}) \\ &= \frac{1}{(k-t)!} \binom{k-j}{t-j} v^{k-t} + o(v^{k-t}).\end{aligned}$$

The leading coefficient is a multiple of $(k-j)^{k-t}$, which is decreasing in j for $0 \leq j \leq t$. This proves the θ_j are distinct as $v \rightarrow \infty$. \square

The proof of Proposition 4.1 also establishes the first part of Theorem 3.2.

Corollary 4.2. *For large v , the least eigenvalue of M is*

$$\theta_t = \binom{v-t}{k-t} + o(v^{k-t}).$$

5. Eigenvalue estimates for M_1

Our focus now shifts to M_1 . To this end, define

$$B = \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{1} & I \end{array} \right],$$

so that $M_1 = MB$ is M with first column replaced by the constant vector $\binom{k}{t} \binom{v-t}{k-t} \mathbf{1}$.

Observe that the eigenvectors of B are precisely those vectors with first coordinate equal to zero.

The column space of each primitive idempotent E_j for the Johnson scheme $J(t, v)$ can be orthogonally decomposed as

$$\langle \mathbf{e}^{(j)} \rangle \oplus \langle \mathbf{e}^{(j)} \rangle^\perp,$$

where $\mathbf{e}^{(j)}$ is a unit vector parallel to the first column of E_j and its complement $\langle \mathbf{e}^{(j)} \rangle^\perp$ is B -invariant.

Let $V = [\mathbf{e}^{(0)} \dots \mathbf{e}^{(t)}]$ and let V_0 be the matrix whose columns are a union of orthonormal bases for the $\langle \mathbf{e}^{(j)} \rangle^\perp$.

Proposition 5.1. *The eigenvalues of M_1 have order v^{k-t} , with $\theta_t/2$ as a lower bound.*

Proof. Let $Q = [V \ V_0]$, an orthogonal matrix. Then

$$Q^\top MBQ = \begin{bmatrix} R & O \\ * & D \end{bmatrix},$$

where $R = V^\top MBV$, a $(t+1) \times (t+1)$ matrix, and D is the $(n-t-1) \times (n-t-1)$ diagonal matrix having eigenvalues θ_j , each with multiplicity $m_j - 1$. It follows that the characteristic polynomial of $M_1 = MB$ factors as

$$\chi_{MB}(x) = \chi_R(x) \prod_{j=1}^t (x - \theta_j)^{m_j - 1}.$$

We recover the original eigenvalues θ_j as all but $t+1$ of the eigenvalues of M_1 . In light of Proposition 4.1, it remains to consider the eigenvalues of R .

Let $\Theta = \text{diag}(\theta_0, \theta_1, \dots, \theta_t)$. By definition of V , we have $MV = V\Theta$. So, since M is symmetric,

$$R = V^\top MBV = (MV)^\top BV = \Theta V^\top BV.$$

It is a routine calculation that

$$V^\top BV = I + \left(\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \binom{v}{t}^{-1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right) \begin{bmatrix} m_0 & m_1 & \dots & m_t \end{bmatrix}. \quad (5.1)$$

The last term on the right of (5.1) is rank one. Put

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \binom{v}{t}^{-1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{m} = \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_t \end{bmatrix}.$$

Recall for column vectors \mathbf{u} and \mathbf{m} the identity

$$\det(I + \mathbf{u}\mathbf{m}^\top) = 1 + \mathbf{u}^\top \mathbf{m}.$$

It follows that the characteristic polynomial of R can be computed rather easily. We have

$$\begin{aligned} \chi_R(x) &= \det(\Theta(I + \mathbf{u}\mathbf{m}^\top) - xI) \\ &= (1 + \mathbf{u}^\top (\Theta - xI)^{-1} \Theta \mathbf{m}) \chi_\Theta(x) \\ &= \left[1 + \frac{\theta_0 m_0}{\theta_0 - x} - n^{-1} \sum_{j=0}^t \frac{\theta_j m_j}{\theta_j - x} \right] \chi_\Theta(x). \end{aligned} \quad (5.2)$$

Although we are not able to explicitly compute the eigenvalues of R in terms of those of Θ , it is sufficient for our purposes to analyze sign changes and obtain an interlacing result. For this purpose, consider the rational function $\psi(x) = \chi_R(x)/\chi_\Theta(x)$. This is the first factor on the right of (5.2).

Near θ_j , $j > 0$, the dominant term in ψ is $-n^{-1}\theta_j m_j/(\theta_j - x)$, which changes from negative to positive as x increases. The opposite is true near θ_0 .

Recall that $\theta_t < \cdots < \theta_1 < \theta_0$, dictating the sign changes of χ_Θ . Finally, observe

$$\begin{aligned} \psi(\theta_t/2) &> 1 + 1 - n^{-1} \sum \frac{\theta_j m_j}{\theta_j - \theta_j/2} \\ &= 2 - 2(m_0 + m_1 + \cdots + m_t)/n = 0. \end{aligned}$$

These various observations are summarized in Table 1. It follows that R has $t + 1$ different real eigenvalues, each exceeding $\frac{1}{2}\theta_t$. The result now follows from Proposition 4.1. \square

odd t (even degree)								
x	$\theta_t/2$	θ_t	θ_{t-1}	\cdots	θ_2	θ_1	θ_0	∞
$\psi(x)$	+	-+	-+	\cdots	-+	-+	+-	+
$\chi_\Theta(x)$	+	+-	-+	\cdots	-+	+-	-+	+
$\chi_R(x)$	+	-	+	\cdots	+	-	-	+

even t (odd degree)								
x	$\theta_t/2$	θ_t	θ_{t-1}	\cdots	θ_2	θ_1	θ_0	∞
$\psi(x)$	+	-+	-+	\cdots	-+	-+	+-	+
$\chi_\Theta(x)$	+	+-	-+	\cdots	+-	-+	+-	-
$\chi_R(x)$	+	-	+	\cdots	-	+	+	-

Table 1: sign changes near eigenvalues of M

6. Applications and concluding remarks

A much weaker version of Theorem 1.3 was considered in [3] for circulant graphs, where a similar approach of considering linear systems was used. A *circulant* has vertices \mathbb{Z}_v and xy is an edge if and only if $x - y$ belongs to some prescribed set $D \subset \mathbb{Z}_v$ of legal distances (where $0 \notin D$ and $-D = D$).

The rational decomposition of circulants into complete graphs enjoys a nice application in statistics. A *balanced sampling plan avoiding adjacent units*, or $\text{BSA}(v, k, r)$, is a weighting of all k -subsets $K \subset \mathbb{Z}_v$ with probabilities p_K such that

- $\sum_K p_K = 1$;
- the first-order probabilities of selecting unit $x \in \mathbb{Z}_v$, namely $p_x = \sum_{K \ni x} p_K$, are constant; and
- the second-order probabilities of selecting x and y , or p_{xy} , are either 0 (if $\text{dist}(x, y) \leq r$ in \mathbb{Z}_v) or otherwise are constant.

Roughly, a $\text{BSA}(v, k, r)$ is used to sample k -subsets from a population of size v with pairwise balance, but so that units in cyclic proximity are never simultaneously sampled. See [3, 6] for further details.

Another series of applications for the decomposition of non-complete t -graphs arises in communication networks. For instance, suppose nodes in a network of size v are scheduled to turn on and off in k -subsets. The

weight assigned to each k -subset represents the length of time it is active in the schedule. Ideally, every t -subset of nodes is simultaneously active for the same amount of time. However, for certain reasons (perhaps control of interference), certain t -subsets should never be on together. This application is admittedly over-simplified, but it remains very valid in spirit. See [2] and the references therein for the use of certain graph decompositions in mobile ad-hoc networks.

Whenever a question involves edge-decomposition of graphs, it is often connected with design theory. A t -(v, k, λ) *design* is a family \mathcal{F} of k -subsets of a v -set V with t -wise coverage $\mathcal{F}^t(T) = \lambda$ for each $T \in \binom{V}{t}$. See [7, 9] for preliminary references on t -designs.

Clearly, a t -(v, k, λ) design is equivalent to a decomposition of K_v^t into K_k^t with nonnegative rational weights whose denominators are divisors of λ . By Lemma 1.2, $K_k^t \leq_{\mathbb{Q}} K_v^t$, and so such a λ trivially exists. But the focus in design theory is to minimize (or at least control) the parameter λ . An early reference along these lines is [10].

The work of R.M. Wilson [8] has asymptotically (in v) settled the existence of 2-(v, k, λ) designs. In other words, for fixed $k \geq 2$ and $\lambda \geq 0$, there exists v_0 such that for $v \geq v_0$, there exists a 2-(v, k, λ) design whenever v satisfies the *necessary conditions*

$$\binom{k}{2} \mid \lambda \binom{v}{2} \quad \text{and} \quad (k-1) \mid \lambda(v-1).$$

For the dense graph analog, one may ask whether there exists a decomposition $K_k \leq G$. Of course there are necessary conditions as above, such as the size of G being divisible by $\binom{k}{2}$. The doctoral thesis of Torbjörn Gustavsson [5], now over 20 years old, appears to have answered this question in the affirmative for densities extremely close to 1. Gustavsson's techniques are highly algorithmic, and rest on some other unpublished work. So, although the result seems believable, some researchers would like to see more details in this direction. By 'clearing denominators', Theorem 1.3 provides reasonable bounds on (integral) decompositions when G has λ -fold edges for some large λ .

Toward more generality, it would be interesting to extend Theorem 1.3 to non-simple t -graphs G under some conditions. We defer further consideration of this question for a separate and careful analysis.

References

- [1] J.M. Carnicer, T.N.T. Goodman, and J.M. Peña, Linear conditions for positive determinants. *Linear Algebra Appl.* 292 (1999), 39–59.
- [2] P.J. Dukes, C.J. Colbourn and V. Syrotiuk, Ternary schedules for energy-limited sensor networks. *IEEE Trans. Inform. Theory.* 53 (2007), 2791–2798.
- [3] P.J. Dukes and A.C.H. Ling, Existence of balanced sampling plans avoiding cyclic distances. *Metrika* 70 (2009), 131–140.
- [4] C.D. Godsil, Notes on association schemes, University of Waterloo, 2005.
- [5] T. Gustavsson, Decompositions of large graphs and digraphs with high minimum degree, Doctoral Dissertation, Department of Mathematics, Stockholm University, 1991.
- [6] J. Stufken, Combinatorial and statistical aspects of sampling plans to avoid the selection of adjacent units. *J. Combin. Inform. System. Sci.* 18 (1993), 81–92.
- [7] J.H. van Lint and R.M. Wilson, *A course in combinatorics*, Cambridge University Press, Cambridge, 1992.
- [8] R.M. Wilson, An existence theory for pairwise balanced designs II: The structure of PBD-closed sets and the existence conjectures. *J. Combin. Theory, Ser. A.* 13 (1972), 246–273.
- [9] R.M. Wilson, On the theory of t designs, in *Enumeration and Design* (Waterloo, Ont., 1982), 19–49, Academic Press, Toronto, 1984.
- [10] R.M. Wilson, The necessary conditions for t -designs are sufficient for something. *Util. Math.* 4 (1973), 207–215.
- [11] R. Yuster, Asymptotically optimal K_k -packings of dense graphs via fractional K_k -decompositions. *J. Combin. Theory Ser. B.* 95 (2005), 1–11.
- [12] R. Yuster, Fractional decompositions of dense hypergraphs. *Bull. Lond. Math. Soc.* 39 (2007), 156–166.